Section 12.2

Check Your Understanding, page 782:

1. Option 1: $\widehat{\text{premium}} = -343 + 8.63(58) = \$157.54$

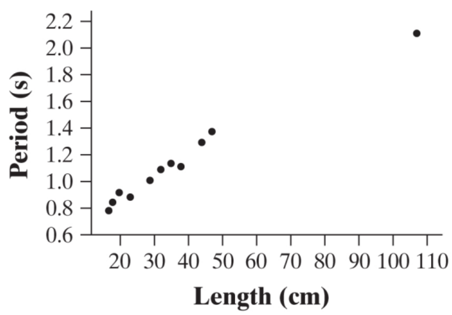
Option 2: $\widehat{\ln(\text{premium})} = -12.98 + 4.416(\ln 58) = 4.9509 \rightarrow \hat{y} = e^{4.9509} = \$141.30$

Option 3: $\widehat{\ln(\text{premium})} = -0.063 + 0.0859(58) = 4.9192 \rightarrow \hat{y} = e^{4.9192} = \$136.89$

2. The exponential model (Option 3) best describes the relationship because the scatterplot showing ln(premium) versus age was the most linear and this model had the most randomly scattered residual plot.

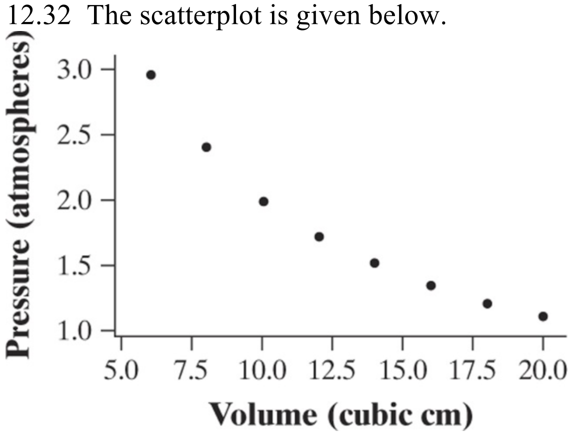
Exercises, page 785:

12.31 The scatterplot is given below.



The scatterplot shows a fairly strong, positive, slightly curved association between length and period with one very unusual point (106.5, 2.115) in the top right corner.
(b) In this case, the class used the square root of x = length to linearize the curved pattern.
(c) In this case, the class used the square of y = period to linearize the curved pattern.

12.32 The scatterplot is given below.



The scatterplot shows a strong, negative, curved association between volume and pressure.
(b) In this case, the students used the reciprocal of x = volume to linearize the curved pattern.
(c) In this case, the students used the reciprocal of y = pressure to linearize the curved pattern.

12.33 (a) For transformation 1: $\hat{y} = -0.08594 + 0.21\sqrt{x}$ where y is the period and x is the length. For transformation 2: $\widehat{y^2} = -0.15465 + 0.0428x$ where y is the period and x is the length.
(b) For transformation 1: $\hat{y} = -0.08594 + 0.21\sqrt{80} = 1.792$ seconds. For transformation 2: $\widehat{y^2} = -0.15465 + 0.0428(80) = 3.269$, so $\hat{y} = \sqrt{3.269} = 1.808$ seconds.

12.34 (a) For transformation 1: $\hat{y} = 0.3677 + 15.8994\left(\frac{1}{x}\right)$ where y is the pressure and x is the volume. For transformation 2: $\widehat{\frac{1}{y}} = 0.1002 + 0.0398x$ where y is the pressure and x is the volume. (b) For transformation 1: $\hat{y} = 0.3677 + 15.8994\left(\frac{1}{17}\right) = 1.303$ atmospheres. For transformation 2: $\widehat{\frac{1}{y}} = 0.1002 + 0.0398(17) = 0.7768$, so $\hat{y} = \frac{1}{0.7768} = 1.287$ atmospheres.

12.35 (a) It is reasonable to use a power model here because the scatterplot of log(period) versus log(length) is roughly linear. Also, the residual plot shows no obvious leftover patterns.
(b) $\widehat{\log y} = -0.73675 + 0.51701\log(x)$ where y is the period and x is the length.

12.36 (a) Even though the residual plot shows a leftover curved pattern, it is reasonable to use a power model here because the scatterplot of log(pressure) versus log (volume) is roughly linear.
(b) $\widehat{\log y} = 1.11116 - 0.81344\log(x)$ where y = pressure and x = volume.

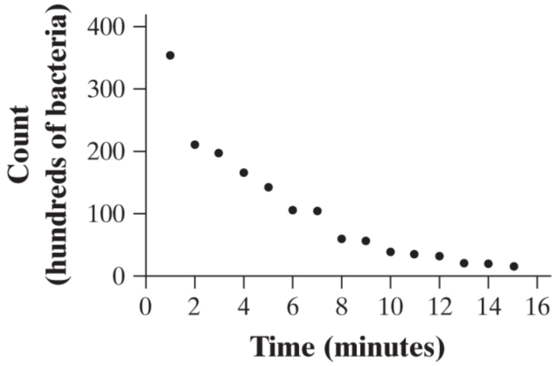
12.37 $\widehat{\log y} = -0.73675 + 0.51701\log(80) = 0.24717$. Thus, $\hat{y} = 10^{0.24717} = 1.77$ seconds.

12.38 $\widehat{\log y} = 1.11116 - 0.81344\log(17) = 0.110264$. Thus, $\hat{y} = 10^{0.110264} = 1.28903$ atmospheres.

12.39 The equation for the regression line is $\widehat{\log y} = 1.01 + 0.72\log x$ where x is body weight in kg and y is brain weight in g. If Sasquatch has a body weight of x = 127 kg, then $\widehat{\log y} = 1.01 + 0.72\log(127) = 2.525$. This means that $\hat{y} = 10^{2.525} = 334.97$ grams is the predicted brain weight of Sasquatch.

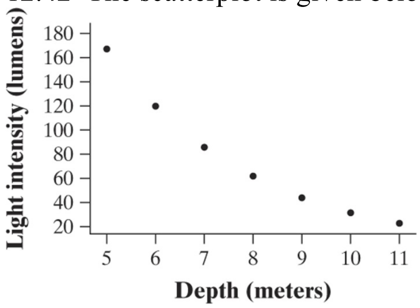
12.40 The equation for the regression line is $\widehat{\ln y} = -2.00 + 2.42\ln x$ where x is the diameter at breast height in cm and y is the aboveground biomass in kg. If a tree is x = 30 cm in diameter, then $\widehat{\ln y} = -2.00 + 2.42\ln(30) = 6.231$. This means that $\hat{y} = e^{6.231} = 508.263$ kg is the predicted total aboveground biomass of the tree.

12.41 (a) The scatterplot is below.



The relationship between bacteria count and time is strong, negative, and curved with a possible outlier in the top left-hand corner.
(b) Because the scatterplot of ln(count) versus time is fairly linear, an exponential model would be reasonable.
(c) $\widehat{\ln y} = 5.97316 - 0.218425x$ where y is the count of surviving bacteria and x is time in minutes.
(d) $\widehat{\ln y} = 5.97316 - 0.218425(17) = 2.26$ so $\hat{y} = e^{2.26} = 9.58$ or 958 bacteria.

12.42 The scatterplot is given below.



The relationship is strong, negative, and slightly curved with no outliers.
(b) Because the scatterplot of ln(intensity) versus depth is fairly linear, an exponential model would be reasonable.
(c) $\widehat{\ln y} = 6.789 - 0.333x$ where y is the light intensity (lumens) and x is the depth (meters).
(d) $\widehat{\ln y} = 6.789 - 0.333(12) = 2.793$ so $\hat{y} = e^{2.793} = 16.33$ lumens.

12.43 (a) The exponential model would work better because the scatterplot of log(height) versus bounce number is roughly linear. A power model would not work as well because the scatterplot of log(height) versus log(bounce) is curved.
(b) $\widehat{\log y} = 0.45374 - 0.11716x$ where y = height in feet and x = bounce number.
(c) $\widehat{\log y} = 0.45374 - 0.11716(7) = -0.36638$ so $\hat{y} = 10^{-0.36638} = 0.43$ feet.
(d) The trend in the residual plot suggests that the residual for $x = 7$ would be positive, meaning that the actual height will be greater that the predicted height. This means that our prediction would be too low.

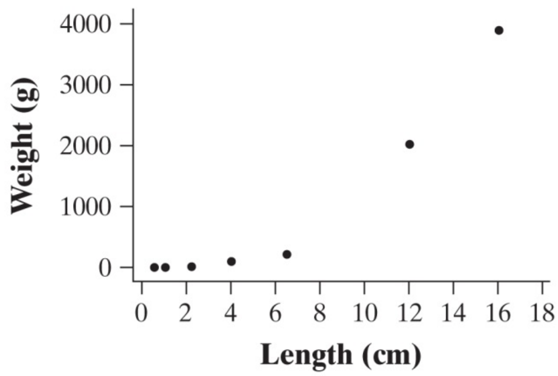
12.44 (a) The power model would work better because the scatterplot of log(abundance) versus log(body mass) is roughly linear. An exponential model would not work as well because the scatterplot of log(abundance) versus body mass is clearly curved.

(b) $\widehat{\log y} = 1.9503 - 1.0481\log x$ where y = abundance (per 10,000 kg of prey) and x = body mass (kg).

(c) $\widehat{\log y} = 1.9503 - 1.0481\log(92.5) = -0.1104$ so $\hat{y} = 10^{-0.1104} = 0.7755$ per 10,000 kg of prey.

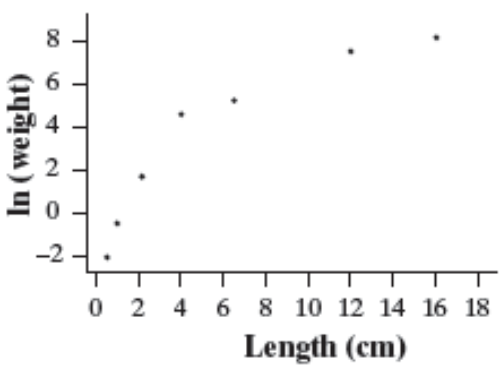
(d) Because there are no leftover patterns in the residual plot, the power model is appropriate for these data.

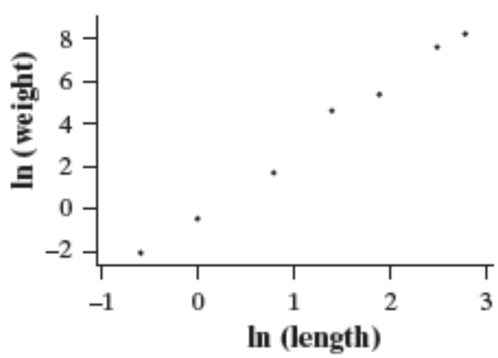
12.45 The scatterplot is below.



There is a strong, positive curved relationship between heart weight and length of left ventricle for mammals.

(b) Two scatterplots are given below.



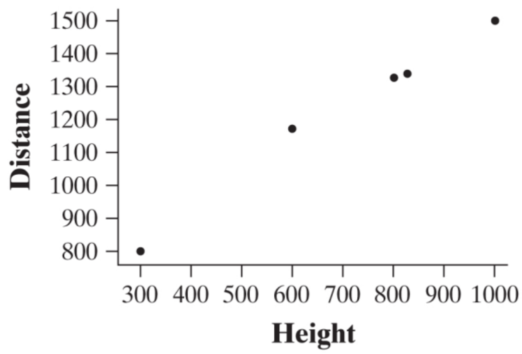


Because the relationship between ln(weight) and ln(length) is roughly linear, heart weight and length seem to follow a power model. An exponential model would not be appropriate because the relationship between ln(weight) and length is clearly curved.

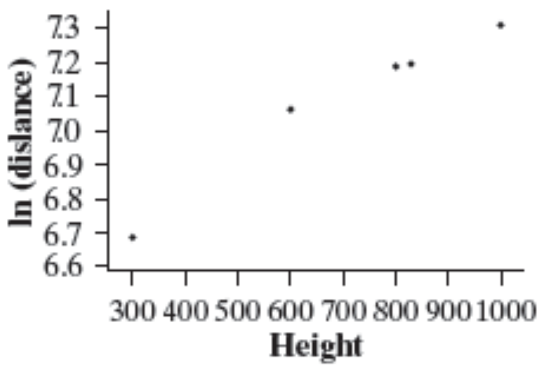
(c) The equation is $\widehat{\ln y} = -0.314 + 3.1387\ln x$ where y is the weight of the heart and x is the length of the cavity of the left ventricle.

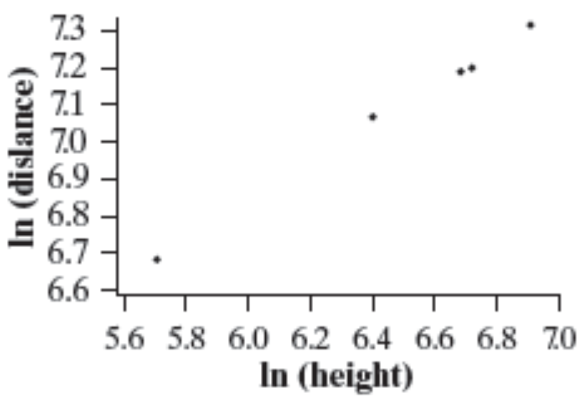
(d) $\widehat{\ln y} = -0.314 + 3.1387\ln(6.8) = 5.703$ so $\hat{y} = e^{5.703} = 299.77$ grams.

12.46 (a) The scatterplot is given below.



There is a strong, positive, slightly curved relationship between height and distance.
(b) Two scatterplots are given below.





Because the relationship between ln(distance) and ln(height) is roughly linear, distance and height seem to follow a power model. An exponential model would not be appropriate because the relationship between ln(distance) and height is still curved.

(c) The equation is $\widehat{\ln y} = 3.7514 + 0.5152 \ln x$ where y is the distance and x is the height.

(d) If the ramp height was 700, $\widehat{\ln y} = 3.7514 + 0.5152 \ln(700) = 7.1265$ and $\hat{y} = e^{7.1265} = 1244.51$ units.

12.47 c

12.48 e

12.49 e

12.50 d

12.51 (a) **Step 1: State the distribution and values of interest.** For Marcela, X = the length of her shower on a randomly selected day follows a Normal distribution with mean 4.5 minutes and standard deviation 0.9 minutes. We want to find $P(3 < X < 6)$. **Step 2: Perform calculations. Show your work.** The standardized scores for the boundary values are $z = \frac{3-4.5}{0.9} = -1.67$ and $z = \frac{6-4.5}{0.9} = 1.67$ From Table A, the proportion of z-scores below $z = -1.67$ is 0.0475 and the proportion of z-scores below 1.67 is 0.9525. Thus, the proportion of z-scores between –1.67 and 1.67 is 0.9525 – 0.0475 = 0.9050. *Using technology:* The command normalcdf(lower: 3, upper: 6, μ: 4.5, σ: 0.9) gives an area of 0.9044. **Step 3: Answer the question.** There is a 0.9044 probability that Marcela's shower lasts between 3 and 6 minutes.

(b) A point is considered to be an outlier if it is more than 1.5*IQR* above Q_3, so we need to find the values of Q_1 and Q_3 for Marcela. **Step 1: State the distribution and values of interest.** For Marcela, X = the length of her shower on a randomly selected day follows a Normal distribution with mean 4.5 minutes and standard deviation 0.9 minutes. Q_1 is the boundary value x with 25% of the distribution to its left and Q_3 is the boundary value with 75% of the distribution to its left. **Step 2: Perform calculations. Show your work.** Look in the body of Table A for the value closest to 0.25. A z-score of –0.67 gives the closest value (0.2514). Solving $-0.67 = \frac{Q_1 - 4.5}{0.9}$ gives Q_1 = 3.897 minutes. Likewise, solving $0.67 = \frac{Q_3 - 4.5}{0.9}$ gives Q_3 = 5.103 minutes. *Using technology:* The command invNorm(area: 0.25, μ: 4.5, σ: 0.9) gives Q_1 = 3.893 minutes and the command invNorm(area: 0.75, μ: 4.5, σ: 0.9) gives Q_3 = 5.107 minutes. Thus, an outlier is any value above 5.107 + 1.5(5.107 – 3.893) = 6.928. **Step 3: Answer the question.** Because 7 > 6.928, a shower of 7 minutes would be considered an outlier for Marcela.

(c) The probability that she takes a 7 minute shower on any given day is $P(X > 7) = P\left(Z > \frac{7-4.5}{0.9}\right) = P(Z > 2.78) = 0.0027$. *Using technology*: normalcdf(lower: 7, upper: 1000, μ: 4.5, σ: 0.9) = 0.0027. Let Y = the number of days that Marcela's shower is 7 minutes or higher. Y is a binomial random variable with $n = 10$ and $p = 0.0027$. $P(Y \geq 2) = 1 - P(Y \leq 1)$ = 1 – binomcdf(trials: 10, p: 0.0027, x value: 1) = 1 – 0.9997 = 0.0003. There is a 0.0003 probability that Marcela's shower time would be 7 minutes or more on at least 2 of the 10 days.

(d) **Step 1: State the distribution and values of interest.** Because the distribution of X is Normal, the sampling distribution of $\bar{x}$ is also Normal. The sampling distribution of $\bar{x}$ has mean $\mu_{\bar{x}} = \mu = 4.5$. Because 10 is less than 10% of days that Marcela showers, the standard deviation is $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{0.9}{\sqrt{10}} = 0.285$. Thus, $\bar{x}$ follows a $N(4.5, 0.285)$ distribution and we want to find P($\bar{x} > 5$). **Step 2: Perform calculations. Show your work.** The standardized score for the boundary value is $z = \dfrac{5-4.5}{0.285} = 1.75$. The desired probability is $P(Z > 1.75) = 1 - 0.9599 = 0.0401$. *Using technology*: normalcdf(lower: 5, upper: 1000, μ : 4.5, σ : 0.285) = 0.0397. **Step 3: Answer the question.** There is a 0.0397 probability that the mean length of Marcela's showers on these 10 days exceeds 5 minutes.

12.52 The sample was taken from only those people who had agreed to participate in Harris polls. It is very likely that people who have agreed ahead of time to participate in polls will have different characteristics from the general population of U.S. adults. So the sample will be representative only of the population of people who have agreed to participate in these polls, not the population of all U.S. adults.

12.53 (a) *State*: We want to estimate the true proportion of all AP teachers attending this workshop who have tattoos at a 95% confidence level. *Plan*: We should use a one-sample z interval for p if the conditions are met. Random: The data come from a random sample. 10%: The sample size (n = 98) is less than 10% of the population of teachers at this workshop (1100). Large Counts: The numbers of successes (23) and failures (75) are at least 10. *Do*: The 95% confidence interval is $0.235 \pm 1.96\sqrt{\dfrac{0.235(0.765)}{98}} = 0.235 \pm 0.084 = (0.151, 0.319)$. *Conclude*: We are 95% confident that the interval from 0.151 to 0.319 captures the true proportion of AP teachers at this workshop who have tattoos.
(b) Yes. Because the value 0.14 is not included in the interval, we have convincing evidence that the true proportion of teachers at the workshop who have a tattoo is not 0.14.
(c) If we had two more successes, this would increase the sample proportion and shift the interval higher. Because the interval was already entirely greater than 0.14, the interval would still not include 0.14. If we had two more failures, the interval will shift to lower values and might include the value 0.14. However, the new interval is $0.23 \pm 1.96\sqrt{\dfrac{0.23(0.77)}{100}} = 0.23 \pm 0.082 =$ (0.148, 0.312), which does not include the value 0.14. So, the answer would not change if we got responses from the 2 nonresponders.

12.54 (a) Answers will vary. A possible response would be that (1) it would be hard to find all of the individual teachers in the sample since they were spread out among 40 different sessions and (2) there may not have been many teachers in a particular subject and in an SRS it is possible that no teachers would be selected from some subjects.
(b) A stratified random sample would allow the statistics teachers to make sure that all subjects were included in appropriate proportions. It would also provide a more precise estimate of the proportion with tattoos if the teachers in each strata are similar to each other and different than teachers in other strata.